THE McELIECE CRYPTOSYSTEM WITH ARRAY CODES

Vedat ŞİAP*

*Department of Mathematics, Faculty of Science and Art, Sakarya University, 54187, Serdivan, Sakarya-TURKEY. vedatsiap@gmail.com

Abstract: Public-key cryptosystems form an important part of cryptography. In these systems, every user has a public and a private key. The public key allows other users to encrypt messages, which can only be decoded using the secret private key. In that way, public-key cryptosystems allow easy and secure communication between all users without the need to actually meet and exchange keys. One such system is the McEliece Public-Key cryptosystem, sometimes also called McEliece Scheme. However, as we live in the information age, coding is used in order to protect or correct the messages in the transferring or the storing processes. So, linear codes are important in the transferring or the storing. Due to richness of their structure array codes which are linear are also an important codes. However, the information is then transferred into the source more securely by increasing the error correction capability with array codes. In this paper, we combine two interesting topics, McEliece cryptosystem and array codes.

Key words: Public-Key cryptosystem, Coding theory, Linear codes, Array codes, McEliece cryptosystem.
AMS Classification: 11T71.

MATRİS KODLAR İLE McELIECE ŞİFRELEME SİSTEMİ

Özet: Açık anahtarlı şifreleme sistemleri kriptografinin önemli bir parçasını oluşturmaktadır. Bu sistemlerde, her kullanıcı açık ve gizli anahtar adını alan iki tür anahtara sahip olup açık anahtar, sadece gizli anahtar kullanılarak şifresi çözülebilen mesajları şifrelemek için diğer kullanıcılara izin vermektedir. Bu şekilde, açık anahtarlı şifreleme sistemleri anahtar değişimi ve bir noktada bağlantıya gerek duymadan bütün kullanıcılar arasında güvenli ve kolay bir iletişime olanak sağlamaktadır. McEliece şeması olarak ta adlandırılan McEliece açık anahtarlı şifreleme sistemleri bu tip şifreleme sistemlerine bir örnek teşkil etmektedir. Bununla birlikte, bilgi çağını yaşadığımız bu günlerde bilginin transferi ya da depolanması aşamasında meydana gelebilecek bilgi zedelenmelerini koruma ve düzeltme amacıyla kodlama kullanılmaktadır. Bu anlamda kullanılan kodlar içinde lineer kodlar önemli bir yer tutmaktadır. Lineer kodlar ailesinden olan matris kodlar zengin bir yapıya sahip olup bu kodlar ile hata düzeltme kabiliyetleri artmakta ve bunun sonucunda bilgi daha güvenilir bir şekilde iletilmektedir. Bu bağlamda, makalede güvenilirliği arttırma adına McEliece şifreleme sistemi inşasında matris kodları göz önüne alınacaktır.

Anahtar kelimeler: Açık anahtarlı şifreleme sistemleri, Kodlama teorisi, Lineer kodlar, Matris kodlar, McEliece şifreleme sistemleri.

I. INTRODUCTION

The McEliece public-key encryption scheme is based on error-correcting codes. The idea behind this scheme is to first select a particular (linear) code for which an efficient decoding algorithm is known, and then to use a trapdoor function to disguise the code as a general linear code. Since the problem of decoding an arbitrary linear code is NP-hard, a description of the original code can serve as the private key, while a description of the transformed code serves as the public key [1].

The McEliece encryption scheme (when used with Goppa codes, as originally proposed by McEliece in 1978) has resisted cryptanalysis to date. It is also notable as being the first public-key encryption scheme to use randomization in the encryption process. Although very efficient, the McEliece encryption scheme has received little attention in practice because of the very large public keys [2], [3].

This paper investigates the application of array codes in cryptography, with special attention to the application in the McEliece cryptosystem. The rest of paper is organized as follows: Section II gives an introduction to cryptography. It explains the terms necessary to understand the rest of the paper and includes linear and array codes. In Section III, we introduce the McEliece cryptosystem. We describe the way the system works. In Section IV, we use array codes to construct a code that can be used within the McEliece cryptosystem.

II. CRYPTOGRAPHIC BACKGROUND

In this section, we present some cryptographic background needed to understand array codes.

In general, we consider words of fixed length n with letters from a finite alphabet Q . Thus words are elements of Q^n . A code is a subset of Q^n and the elements of the code are called codewords. The natural number n is the length of the code.

An important class of codes is linear codes. This will be the only class of codes considered in this paper.

2.1 Linear Codes

From now on let the alphabet Q be a finite field F_q , so $Q^n = F_q^n$ is a vector space.

Definition 2.1. [4] (Hamming distance, weight). To give the difference of two codewords a precise meaning the (Hamming) distance between two words is introduced. Let $x, y \in F_q^n$, then

$$d(x,y) = \left|\{i : x_i \neq y_i\}\right| . \qquad (2.1)$$

The (Hamming) weight of a codeword is the number of non-zero entries and therefore the distance from the zero vector:

$$w(x) = \left|\{i : x_i \neq 0\}\right| = d(x,0) . \qquad (2.2)$$

Definition 2.2. [4] A linear code C of dimension k is a k-dimensional linear subspace of F_q^n and is often called an $[n,k]$ code.

The third important parameter of a code C , besides the length and dimension, is the minimum distance between its codewords.

Definition 2.3. [4] The minimum Hamming distance d of a linear code is

$$d = \min_{u \neq v} d(u,v) = \min_{u \neq 0} w(u) . \qquad (2.3)$$

It is often called the minimum distance or simply the distance of the code; any two codewords differ in at least d places.

A code of dimension k , length n and minimum distance d is often called an $[n,k,d]$ code.

Two types of matrices play an important role for linear codes: generator and (parity) check matrices. They are defined as follows.

Definition 2.4. [4] If the encoding $\varepsilon : F_q^k \to F_q^n$ from message m to codeword c is done by the matrix multiplication

$$c = \varepsilon(m) = mG , \qquad (2.4)$$

where G is a $k \times n$ matrix with entries in F_q , then G is called generator matrix of the code. The rows of G form a basis of C .

Definition 2.5. [4] A parity check matrix of a linear $[n,k]$ code C is a $(n-k) \times n$ matrix H , such that

$$C = \left\{x \in F_q^n : xH^T = 0\right\} . \qquad (2.5)$$

Thus the rows of a check matrix generate the orthogonal complement of C .

Example 2.1. The binary code $C = \{0000, 0101, 1110, 1011\}$ can be defined by a generator matrix G , where

$$G=\begin{pmatrix}1&1&1&0\\0&1&0&1\end{pmatrix}. \tag{2.5}$$

It can also be defined by a check matrix H with

$$H=\begin{pmatrix}1&0&1&0\\1&1&0&1\end{pmatrix}. \tag{2.6}$$

In many cases we transmit the encoded message. Because of the channel noise the received word may contain some errors, so we want to be able to at least detect or better correct these errors. Usually we do this by choosing the codeword which is closest (with respect to the Hamming metric) to the received word to minimize the probability of making a mistake.

Lemma 2.1. [4] For an $[n,k,d]$-code C the spheres

$$S_c=\left\{x\in F_q^n : d(x,c)\le\left\lfloor\frac{1}{2}(d-1)\right\rfloor\right\},\quad c\in C, \tag{2.7}$$

do not overlap, so every received word in S_c will be corrected to c. Hence this code corrects up to $\left\lfloor\frac{1}{2}(d-1)\right\rfloor$ errors.

Proof. Assume two of the spheres overlap, i.e. they both contain a point $x\in F_q^n$. Then the distance between the two centers of the spheres is not greater than twice the distance to x, thus not greater than $(d-1)$. This contradicts the assumption that C is a code with minimum distance d.

For any given linear code we can construct its dual code.

Definition 2.6. [4] If C is an $[n,k]$ linear code over F_q, its dual or orthogonal code $C^{\perp}$ is the set of vectors which are orthogonal to all codewords of C:

$$C^{\perp}=\left\{u\in F_q^n : uv^t=0 \text{ for all } v\in C\right\}. \tag{2.8}$$

Now, we define matrix space and array codes as follows:

Definition 2.7. [5] Let F_q be a finite field with q elements. Let $Mat_{m\times n}(F_q)$ denote the set of matrices with entries in F_q. Then, $Mat_{m\times n}(F_q)$ is a vector space over F_q.

Definition 2.8. [5] An array code is a subset of $Mat_{m\times n}(F_q)$ and a linear array code is an F_q-linear subspace of $Mat_{m\times n}(F_q)$.

Example 2.2. The elements of the linear array code $C\subset Mat_{2\times 2}(F_3)$ spanned by $v_1=\begin{pmatrix}1&0\\0&1\end{pmatrix}$ and $v_2=\begin{pmatrix}1&1\\0&1\end{pmatrix}$ are as follows:

$$C=\left\{\begin{matrix}\begin{pmatrix}0&0\\0&0\end{pmatrix},\begin{pmatrix}1&0\\0&1\end{pmatrix},\begin{pmatrix}1&1\\0&1\end{pmatrix},\begin{pmatrix}2&1\\0&2\end{pmatrix},\begin{pmatrix}2&0\\0&2\end{pmatrix},\\ \begin{pmatrix}2&2\\0&2\end{pmatrix},\begin{pmatrix}0&1\\0&0\end{pmatrix},\begin{pmatrix}1&2\\0&1\end{pmatrix},\begin{pmatrix}0&2\\2&0\end{pmatrix}\end{matrix}\right\}.$$

Note that the space $Mat_{m\times n}(F_q)$ is identifiable with the space F_q^{mn}. Every matrix in $Mat_{m\times n}(F_q)$ can be represented as a $1\times mn$ vector by writing the first row of matrix followed by second row and so on. Similarly, every vector in F_q^{mn} can be represented as an $m\times s$ matrix in $Mat_{m\times n}(F_q)$ by separating the co-ordinates of the vector into m groups of s-coordinates [5].

Definition 2.9. [5] The mapping with parameters m and n, denoted by $M_{m,n}(.)$, is the one that maps the vector $v=(v_1,v_2,....,v_{mn})$ into matrix $A=\begin{pmatrix}a_{0,0}&\cdots&a_{0,n-1}\\ .&.&.\\ a_{m-1,0}&\cdots&a_{m-1,n-1}\end{pmatrix}_{m\times n}$ so that $a_{i,j}=v_{in+j+1}$, for $i=0,1,...,m-1$, $j=0,1,...,n-1$.

Example 2.3. Let $m=3$, $n=4$ and $e=(0,0,1,0,4,0,0,0,0,0,0,3)\in F_5^{12}$. Then,

$$A=\begin{pmatrix}0&0&1&0\\4&0&0&0\\0&0&0&3\end{pmatrix}.$$

Definition 2.10. [6] Given an array $(a_{i,j})_{\substack{0\le i\le m-1\\0\le j\le n-1}}$, its horizontal syndrome $h=(h_0,h_1,...,h_{m-1})$ and $v=(v_0,v_1,...,v_{n-1})$ are defined by

$$h_i = \sum_{l=0}^{n-1} a_{i,l}, \quad 0 \le i \le m-1,$$
$$v_i = \sum_{l=0}^{m-1} a_{l,j}, \quad 0 \le j \le n-1.$$

Clearly, an array is in $Mat_{m\times n}$ if and only if both the horizontal and the vertical syndromes are equal to zero. As an example, consider the array

$$A = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

Its horizontal syndrome is $h = 101$, while its vertical syndrome is $v = 0100001$. It follows that this array is not in $Mat_{3\times 7}$.

The related work about array codes can be found in [7].

III. THE McELIECE CRYPTOSYSTEM

Let C be an $[n,k]$ linear code with a fast decoding algorithm that can correct up to t errors. Let G be a generator matrix for C. To create the disguise, let S be a random $k\times k$ invertible matrix (also called the scrambler) and let P be a random $n\times n$ permutation matrix. The matrix

$$\widehat{G} = SGP \qquad (2.9)$$

is made public while S, G and P form the private key.

Encryption: Represent the message as a string m of length k, choose a random error vector e of weight at most t and compute the cipher text $c = m\widehat{G} + e$.

Decryption: To recover the plaintext m from c, we compute $\widehat{c} = cP^{-1}$, use the decoding algorithm for the code generated by G to decode $\widehat{c}$ to $\widehat{m}$ and compute $m = \widehat{m}S^{-1}$.

Proof that decryption works. Since

$$\widehat{c} = cP^{-1} = \left(m\widehat{G} + z\right)P^{-1} = \left(mSGP + z\right)P^{-1} = \left(mS\right)G + zP^{-1}$$

and zP^{-1} has weight at most t, the decoding algorithm for the code generated by G corrects $\widehat{c}$ to $\widehat{m} = mS$. Finally, $\widehat{mS}^{-1} = m$ and, hence, decryption works [2].

3.1. Practicality of the McEliece Scheme

As pointed out by Rao and Nam, the McEliece scheme requires rather large block length. They suggested $n = 1024$, but today this is not enough anymore, so $n = 2048$ should be chosen. Therefore this scheme produces too much computational overhead for encryption and decryption for most practical applications [8].

IV. A McELIECE CRYPTOSYSTEM USING ARRAY CODES

In this section, we will discuss how to construct a McEliece cryptosystem using array codes [7].

Example 4.1. Let C be a code over F_5. Consider the matrix

$$G = \begin{pmatrix} 1 & 4 & 0 & 0 & 4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 4 & 0 & 4 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 4 & 0 & 0 & 4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 4 & 1 & 0 & 0 & 1 & 4 & 0 & 0 \\ 0 & 4 & 1 & 0 & 0 & 1 & 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 1 & 0 & 1 & 0 & 4 & 0 & 0 & 0 & 0 \end{pmatrix}$$

which is the generator matrix for the code. Suppose Alice wishes to send a message

$$m = \left(3,2,0,1,4,2\right)$$

to Bob. In order to do so, Bob must create an invertible matrix S and a random permutation matrix P that he will keep secret. If Bob chooses

$$S = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

and

$$P=\begin{pmatrix}1&0&0&0&0&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0&0&0&0&0\\0&0&1&0&0&0&0&0&0&0&0&0\\0&0&0&1&0&0&0&0&0&0&0&0\\0&0&0&0&1&0&0&0&0&0&0&0\\0&0&0&0&0&1&0&0&0&0&0&0\\0&0&0&0&0&0&1&0&0&0&0&0\\0&0&0&0&0&0&0&1&0&0&0&0\\0&0&0&0&0&0&0&0&1&0&0&0\\0&0&0&0&0&0&0&0&0&1&0&0\\0&0&0&0&0&0&0&0&0&0&0&1\\ \end{pmatrix}.$$

Using these, Bob generates the public encryption matrix

$$\widehat{G}=\begin{pmatrix}0&0&0&0&0&1&4&0&0&4&1&0\\1&4&0&0&4&1&0&0&0&0&0&0\\0&0&0&0&0&1&0&4&0&4&0&1\\0&4&1&0&0&1&4&0&0&0&0&0\\0&0&0&0&4&1&0&0&1&4&0&0\\0&4&0&1&0&1&0&4&0&0&0&0\end{pmatrix}.$$

In order to encrypt, Alice generates her own random error vector e and calculates the cipher text $c=m\widehat{G}+e$. In the case of the array code the error vector has weight 1. Suppose Alice chooses

$$e=(0,1,0,0,0,0,0,0,0,0,0,0).$$

Thus the received word is

$$\begin{aligned}c&=m\widehat{G}+e=(2,0,1,2,4,2,1,3,4,3,3,0)+(0,1,0,0,0,0,0,0,0,0,0,0)\\&=(2,1,1,2,4,2,1,3,4,3,3,0).\end{aligned}$$

The decoding procedure need only c into the matrix $M_{3,4}(c)$ from Definition 2.9 and find the errors by checking these rows and columns, and remove the errors via horizontal and vertical syndromes. The procedure of finding errors is demonstrated in the following.

From c, we know that $M_{3,4}(c)=\begin{pmatrix}2&1&1&2\\4&2&1&3\\4&3&3&0\end{pmatrix}$. Its horizontal and vertical syndromes are $h=100$ and $v=0100$, respectively. That is, by checking bits, we find that the first row contains one error 1 and the second column contains one error 1. It is clear that the error vector is $(0,1,0,0,0,0,0,0,0,0,0,0)$. Therefore, the errors can be found and removed.

V. CONCLUSION

Array codes have the potential to improve the use of the McEliece cryptosystem.

REFERENCES

1. Stinson, D.R., *Cryptography theory and practice,* CRC Press LLC, USA (1995).
2. McEliece, R.J., *A public-key cryptosystem based on algebraic coding theory,* DSN Progress Report 42-44, 114-116 (1978).
3. Golay, M.J.E., *Notes on digital coding,* Proc. I.R.E., 37, 657 (1949).
4. Roman, *Coding and Information Theory,* Graduate Text in Mathematics, Springer Verlag (1992).
5. Sapna, J., *Campopiano-type bounds in non-Hamming array coding,* Linear Algebra and its Applications, 420, 135-159 (2007).
6. Pless, V.S., Hufmann, W.C., *Handbook of coding theory,* Elsevier B.V., The Netherlands (1998).
7. Şiap, V., *Matris Kodlar ile McEliece Şifreleme Sistemi,* Yüksek Lisans, Sakarya Üniversitesi, 55-73 (2008).
8. Rao, T.R.N., Nam, K.H., *Private-key algebraic-code encryptions,* IEEE Trans on Inform Theory, 35 (1989).